Volume 7 Number 8 1979 Nucleic Acids Research

Regions of the polyoma genome coding for T antigens

T.Hunter, M.A.Hutchinson, W.Eckhart, T.Friedmann*, A.Esty*, P.LaPorte* and P.Deininger*

Tumor Virology Laboratory, The Salk Institute, P.O. Box 85800, San Diego, CA 92138, and
*Department of Pediatrics, School of Medicine, University of California, San Diego, La Jolla, CA 92093, USA

Received 29 August 1979

ABSTRACT

The early region of the polyoma genome encodes three T antigens. We have analyzed the organization of the coding regions for the T antigens, using the nucleotide sequence of polyoma DNA and peptides derived from purified, radio-labeled T antigens, separated by two-dimensional electrophoresis and chromatography. We compared the peptides, predicted from the nucleotide sequence of the DNA, with those derived from the purified T antigens. We also compared chemically synthesized peptides, predicted from the DNA sequence, with observed peptides. The results show that the three polyoma T antigens are encoded in overlapping regions of the viral DNA, translated, in part, in two different reading frames.

INTRODUCTION

The proteins required for cell transformation by the DNA tumor virus, polyoma, are encoded in the early region of the viral genome. Such proteins can be identified by immunoprecipitation of infected cell extracts, using serum from tumor-bearing animals, and are referred to as T antigens. In vitro translation of virus-specific RNA shows that polyoma encodes three T antigens (1). These are designated large, medium and small, and have molecular weights of approximately 90,000 (90K), 60,000 (60K) and 22,000 (22K), respectively (2-6).

The involvement of T antigens in cell transformation is shown by the properties of polyoma mutants defective in transformation. The temperature-sensitive tsA mutants render the large T antigen thermolabile at the non-permissive temperature (2,5,6) and fail to transform infected cells at the nonpermissive temperature. The host range non-transforming (hr-t) mutants affect the small and medium T antigens, but not the large T antigen, and also fail to transform infected cells (3,5,6). The tsA and the hr-t mutants complement each other in mixed infection for transformation, showing that at least two functions required for transformation are encoded in the early region of the polyoma genome (7,8).

© Information Retrieval Limited 1 Falconberg Court London W1V 5FG England

The arrangement of coding regions for the polyoma T antigens has been inferred from the location of mutations affecting the T antigens, and from an analysis of the relationships among tryptic peptides derived from isolated T antigens (6,9). All three T antigens share a common N-terminal region encoded between 74 and 79 map units on the physical map of the viral genome; the small and medium T antigens share amino acid sequences not present in the large T antigen, encoded between 79 and 85 map units; the large T antigen has unique amino acid sequences encoded between 86 and 26 map units (clockwise), and the medium T antigen has unique sequences encoded between 86 and 99 map units.

There is extensive amino acid and nucleotide sequence homology between the early regions of polyoma and SV40 (10-12). The early region of the SV40 genome encodes a large and a small T antigen, 81,600 and 20,500 daltons molecular weight, respectively (13,14), but no virus-coded T antigen corresponding to the polyoma medium T has been identified.

To define more precisely the organization of the coding region for T antigens on the polyoma genome, we have carried out an extensive analysis of the methionine and cysteine-containing peptides produced from isolated T antigens by digestion with various proteases, and have compared them with the peptides predicted from the nucleotide sequence of polyoma DNA.

MATERIALS AND METHODS

Analysis of T antigen peptides. Polyoma T antigens were immunoprecipitated and analyzed by SDS polyacrylamide gel electrophoresis as described (6). For peptide analysis the T antigens were labeled with ^{35}S methionine (100 μCi/ml, Amersham/Searle, > 500 Ci/mmole) or ^{35}S cystine (200 μCi/ml, New England Nuclear, >500 Ci/mmole) for 5 hours during the late phase of lytic infection. The T antigens were immunoprecipitated, isolated and prepared for digestion as described (6). The labeled T antigens were digested with TPCK-treated trypsin (Worthington) as described (6), and further digested with 10 μg of chymotrypsin (Worthington) or S. aureus V8 protease (Miles) in 0.01 M NH_4HCO_3 buffer, pH 7.8, overnight at 25° C, followed by 1 hr at 37° C. The peptides were separated by electrophoresis at pH 4.7, followed by ascending chromatography on thin layer plates (6). Detection of radioactive spots by fluorography was enhanced by the use of 2 methyl-naphthalene.

Prediction of peptide mobilities. All 20 individual amino acids and methionine sulfone and cysteic acid (the end states of performic acid oxidation) were analyzed by electrophoresis and chromatography under conditions identical to those used for separation of the T antigen peptide mixtures and the distances migrated from the origin were measured. Despite their electrical neutrality

at pH 4.7 the amino acids, alanine, asparagine, glutamine, glycine, isoleucine, leucine, methionine (and methionine sulfone), phenylalanine, proline, serine, threonine, tryptophan, tyrosine and valine, all migrated together, due to endosmosis, towards the cathode, about one-third of the distance moved by lysine. The point to which the neutral amino acids migrated was taken as the position to which uncharged molecules would move and was used in calculating mobilities. Using the known pK_a's lysine has, by definition, a net charge of +1 at pH 4.7. Net charges were calculated from the distances moved relative to lysine. Arginine and histidine both had experimentally determined charges of +1, while cysteic acid, aspartic acid and glutamic acid had net charges of -1, -0.67 and -0.46, respectively. The overall net charge on a peptide was calculated assuming that the charges on the amino group at the N-terminus and the carboxyl group at the C-terminus cancel out. For a given peptide there is an inverse relationship between the logarithm of the (mass/charge) ratio and the logarithm of the mobility, as described by Offord (15). We used this relationship to determine an approximate electrophoretic mobility for each theoretical peptide, using the mobilities of known peptides as standards.

In the chromatographic dimension the R_f's of the individual amino acids were determined relative to the mobility of ε DNP-lysine and were as follows: alanine (0.35), arginine (0.15), aspartic acid (0.14), cysteic acid (0.08), glutamic acid (0.22), glutamine (0.20), glycine (0.23), histidine (0.19), lysine (0.14), methionine (0.65), methionine sulfone (0.25), phenylalanine (0.77), proline (0.40), serine (0.25), threonine (0.34), tryptophan (0.69), tyrosine (0.64) and valine (0.56). It can be seen that the more hydrophobic amino acids have higher R_f values in this chromatographic system. To calculate a predicted chromatographic mobility for a peptide, the sum of the mobilities of all the constituent amino acids was divided by the number of residues. In the cases where we have analyzed peptides of known sequence the predicted two-dimensional mobilities were reasonably accurate but by no means exact.

Nucleotide sequence ananlysis. The nucleotide sequence of polyoma DNA was analyzed as described (11) using methods modified from those of Maxam and Gilbert (16) and the dideoxynucleotide sequencing procedure of Sanger (17) as modified by Smith (18). Sequences were stored and analyzed in a Burroughs 6700 computer, using the program of Staden (19).

RESULTS

Strategy for comparing predicted and observed T antigen peptides. To compare the peptides of isolated T antigens with the peptides predicted from the

nucleotide sequence of polyoma DNA we have analyzed the T antigens in the following ways:

1) By digesting methionine-labeled proteins with trypsin (cleaving after lysine or arginine), followed by further digestion with chymotrypsin (cleaving after hydrophobic amino acids, especially phenylalanine, tyrosine and tryptophan) or with S. aureus V8 protease (cleaving after glutamic acid under the conditions used).

2) By digesting cysteine-labeled proteins as above, and analyzing mixtures of the methionine and cysteine-labeled digests. The number of peptides produced by proteolytic cleavage, and the presence or absence of amino acids sensitive to enzymatic attack, was compared to predictions of peptide number and composition derived from the nucleotide sequence of the viral DNA.

3) By calculating the expected electrophoretic and chromatographic mobilities of the predicted peptides and comparing them to the properties of the observed peptides.

4) By comparing the migration of chemically-synthesized peptides, predicted from the DNA sequence, to the migration of observed peptides.

Methionine-containing tryptic peptides. Figure 1 shows the methionine-containing proteins immunoprecipitated from an extract of polyoma-infected mouse 3T6 cells. The large, medium and small T antigens are designated L, M and S, respectively. Other bands correspond to the major capsid protein, VP1, and proteins which are most likely coded for by cellular genes (1,6).

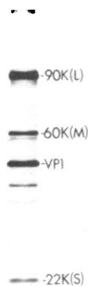

Fig. 1 - Polyoma T Antigens. Polyoma-infected 3T6 cells, labeled with ^{35}S methionine, were extracted during the late phase of lytic infection. The extracts were immunoprecipitated with rat antitumor serum and analyzed by gel electrophoresis.

The methionine-labeled T antigens were isolated from a preparative gel, digested with trypsin, and analyzed by two-dimensional electrophoresis and chromatography. Figure 2 shows the radioactive peptides from the isolated T antigens. All three T antigens share some peptides in common (labeled 1-4); the small and medium T share some peptides (A-E) not present in the large T antigen, and the medium T antigen contains several unique peptides (U-Z). Previously it was proposed that the peptides common to all three T antigens were encoded between 74 and 79 map units; the peptides common to the small and medium T antigens between 79 and 85 map units; and the peptides unique to the medium

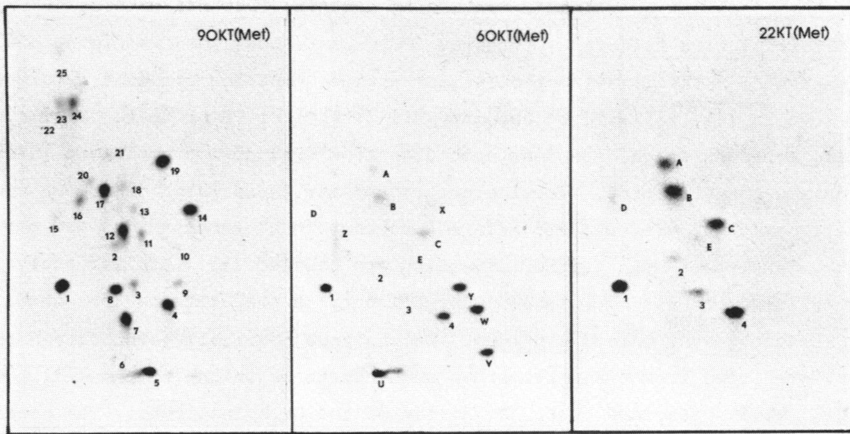

Fig. 2 - Methionine-containing tryptic peptides of the three polyoma T antigens. Polyoma T antigens, labeled with ^{35}S methionine, were isolated from a preparative SDS polyacrylamide gel and digested with trypsin. The resulting peptides were separated by two-dimensional electrophoresis and chromatography.

T antigen between 86 and 99 map units (6,9).

Cysteine-containing tryptic peptides. Figure 3 shows the two-dimensional separation of tryptic digests of the isolated, cysteine-labeled, T antigens. There is no cysteine-containing peptide common to all three T antigens. The small and medium T antigens share peptides B and D, containing both cysteine

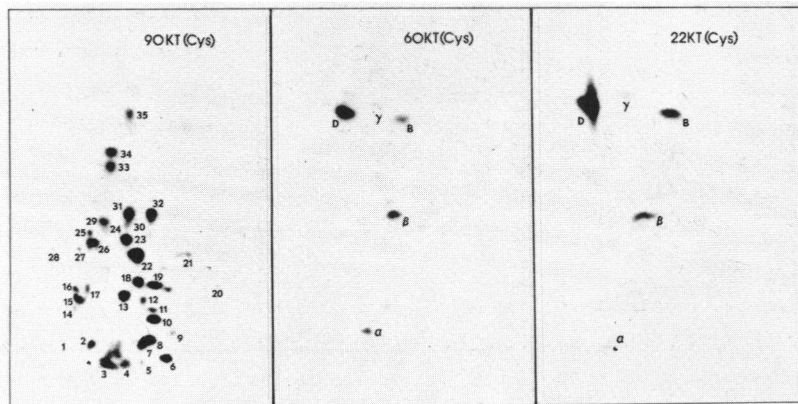

Fig. 3 - Cysteine-containing tryptic peptides of the three polyoma T antigens. Polyoma T antigens, labeled with ^{35}S cysteine, were isolated, digested with trypsin, and separated by two-dimensional electrophoresis and chromatography.

and methionine, and α, β and γ, containing cysteine, but not methionine.

Digestion with different proteases. Figure 4 shows the two-dimensional separation of methionine-labeled large T antigen peptides produced by digestion with trypsin (T), followed by chymotrypsin (T + C) or V8 protease (T + S). Tryptic peptides which disappear upon digestion with chymotrypsin are inferred to contain phenylalanine, tyrosine or tryptophan; those which disappear upon digestion with V8 protease are inferred to contain glutamic acid. New peptides which appear upon further digestion are labeled (x). Similar analyses were performed on cysteine-labeled large T antigen preparations (not shown).

Chemically-synthesized peptides. We analyzed chemically-synthesized peptides, predicted from the polyoma DNA sequence to be unique to the medium T antigen, kindly provided by Dr. J. Rivier of the Salk Institute. The peptides (Met-Arg and Ala-His-Ser-Met-Gln-Arg), would be encoded at 96 (nucleotides 1360-1365) and 96.5 (nucleotides 1390-1407) map units, respectively, in a region of the genome suspected to encode both medium and large T antigens, but translated in two different reading frames. Figure 5 shows that Met-Arg comigrates with the methionine-containing peptide, W, and Ala-His-Ser-Met-Gln-Arg comigrates with V, both of which are unique to the medium T antigen. The chances of the tripeptide $^{\text{Lys}}_{\text{Arg}}$-Met-Arg occurring at random in a protein sequence is 1 in 4000. This is reasonably infrequent and a computer search of all the protein sequences available in 1977 for such a tripeptide conducted by Dr. R.

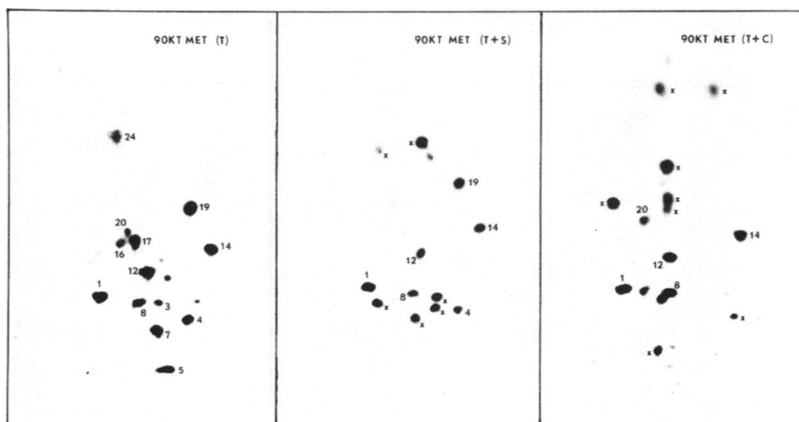

Fig. 4 - Sensitivity of polyoma large T antigen tryptic peptides to chymotrypsin and V8 protease. Polyoma large T antigen, labeled with ^{35}S methionine, was digested with trypsin alone (T), or trypsin followed by chymotrypsin (T + C) or V8 protease (T + S). The mixtures were separated by two-dimensional electrophoresis and chromatography.

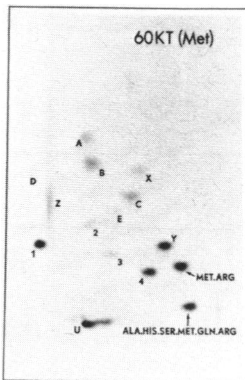

Fig. 5 - Comigration of chemically-synthesized peptides with methionine-containing tryptic peptides of the polyoma medium T antigen. Isolated polyoma medium T antigen, labeled with ^{35}S methionine, was digested with trypsin. Two chemically-synthesized peptides were added, and the mixture was separated by two-dimensional electrophoresis and chromatography.

Burgus at the Salk Institute yielded only four instances. The chances of the heptapeptide $^{\text{Lys}}_{\text{Arg}}$-Ala-His-Ser-Met-Gln-Arg occurring at random is 1 in 6×10^8. The chances of both peptides being found in a protein other than the medium T antigen are clearly negligible. Therefore, this experiment unequivocally establishes the reading frame used for the medium T antigen in this region. We established the reading frame used for the large T antigen at 12.6 map units (nucleotides 2241-2249) by showing that peptide 14 (Figs. 2 and 4) which does not contain cysteine and which is insensitive to cleavage by chymotrypsin or V8 protease, comigrates with the peptide Met-Leu-Lys (see Fig. 7).

Comparison with the nucleotide sequence of polyoma DNA. By comparing the observed peptides to the peptides predicted by the nucleotide sequence of polyoma DNA it is possible to identify with reasonable certainty the coding region for most of the methionine-containing tryptic peptides of the small and medium T antigens described previously, including those shared with the large T antigen. The predicted amino acid sequence for large T antigen has been reported previously (11). Figure 6 shows the nucleotide sequence of polyoma DNA between 73.8 and 99.3 map units [nucleotides 182 to 1560, numbered as described previously (11)] and the predicted amino acid sequence of the presumed coding regions for the small and medium T antigens. The coding region for the C terminal portion of the medium T antigen is translated in a different reading frame from the N-terminal portion; hence the shift in reading frame at position 826. It is not known whether a splice in the mRNA occurs precisely at this position, but the sequence at this site has homology with other known splice sites (11). In addition the structural analysis of

```
          ——1——
182
          Met Asp Arg Val Leu Ser Arg Ala Asp Lys Glu Arg Leu Leu Glu Leu Leu Lys
ACC ATC   ATG GAT AGA GTT CTG AGC AGA GCT GAC AAA GAA AGG CTG CTA GAA CTT CTA AAA

                                                   ————4————
242
Leu Pro Arg Gln Leu Trp Gly Asp Phe Gly Arg Met Gln Gln Ala Tyr Lys Gln Gln Ser
CTT CCC AGA CAA CTA TGG GGG GAT TTT GGA AGA ATG CAG CAG GCA TAT AAG CAG CAG TCA

302                          ——————————————————————3——————
Leu Leu Leu His Pro Asp Lys Gly Gly Ser His Ala Leu Met Gln Glu Leu Asn Ser Leu
CTG CTA CTG CAC CCA GAC AAA GGT GGA AGC CAT GCC TTA ATG CAG GAA TTG AAC AGT CTC

362 ——————————                                ————————————C—————
Trp Gly Thr Phe Lys Thr Glu Val Tyr Asn Leu Arg Met Asn Leu Gly Gly Thr Gly Phe
TGG GGA ACA TTT AAA ACT GAA GTA TAC AAT CTG AGA ATG AAT CTA GGA GGA ACC GGC TTC

422 ————
Gln Val Arg Arg Leu His Ala Asp Gly Trp Asn Leu Ser Thr Lys Asp Thr Phe Gly Asp
CAG GTA AGA AGG CTA CAT GCG GAT GGG TGG AAT CTA AGT ACC AAA GAC ACC TTT GGT GAT

482                             ——————————B—————————
Arg Tyr Tyr Gln Arg Phe Cys Arg Met Pro Leu Thr Cys Leu Val Asn Val Lys Tyr Ser
AGA TAC TAC CAG CGG TTC TGC AGA ATG CCT CTT ACC TGC CTA GTA AAT GTT AAA TAC AGC

542
Ser Cys Ser Cys Ile Leu Cys Leu Leu Arg Lys Gln His Arg Glu Leu Lys Asp Lys Cys
TCA TGT AGT TGT ATA TTA TGC CTG CTT AGA AAG CAA CAT AGA GAG CTC AAA GAC AAA TGT

602             ———————————————————————————————D——————————————
Asp Ala Arg Cys Leu Val Leu Gly Glu Cys Phe Cys Leu Glu Cys Tyr Met Gln Trp Phe
GAT GCC AGG TGC CTA GTA CTT GGA GAA TGT TTT TGT CTT GAA TGT TAC ATG CAA TGG TTT

662 ————————————————   ——————————————————————————————A———————
Gly Thr Pro Thr Arg Asp Val Leu Asn Leu Tyr Ala Asp Phe Ile Ala Ser Met Pro Ile
GGA ACA CCA ACC CGA GAT GTG CTG AAC CTG TAT GCA GAC TTC ATT GCA AGC ATG CCT ATA

722 ——————————————————————————————————————————————— *** Val Ser Arg Gly Arg
Asp Trp Leu Asp Leu Asp Val His Ser Val Tyr Asn Pro Ser Lys Tyr Gln Glu Gly Gly
GAC TGG CTG GAC CTG GAT GTG CAC AGC GTG TAT AAT CCA AGT AAG TAT CAA GAG GGC GGG

782
Val Gly Ile Tyr Gly Leu Tyr Ser Tyr Arg Ala Leu Pro Leu Glu
Trp Val Phe Thr Ala Tyr Ile Leu Thr Gly Leu Ser Pro ***     Arg Arg Ser Glu Glu
TGG GTA TTT ACG GCC TAT ATT CTT ACA GGG CTC TCC CCC TAG  AA CGG CGG AGC GAG GAA

841         ————————————————————————————————————————————————U———
Leu Arg Arg Ala Ala Thr Val His Tyr Thr Met Thr Thr Gly His Ser Ala Met Glu Ala
CTG AGG AGA GCG GCC ACA GTC CAC TAC ACG ATG ACT ACT GGT CAT TCA GCT ATG GAA GCA

901 ————————————————————————————————————————————————————————————————
Ser Thr Ser Gln Gly Asn Gly Met Ile Ser Ser Glu Ser Gly Thr Pro Ala Thr Ser Arg
AGT ACT TCA CAA GGG AAT GGA ATG ATT TCT TCA GAA AGT GGG ACC CCA GCT ACC AGT CGC

961             ———————————————Y————————————————
Arg Leu Arg Leu Pro Ser Leu Leu Ser Asn Pro Thr Tyr Ser Val Met Arg Ser His Ser
CGC CTA AGA CTG CCG AGT CTT CTG AGC AAC CCG ACC TAT TCT GTT ATG AGG AGC CAC TCC

1021                    ————————————————————————————————————————*———
Tyr Pro Pro Thr Arg Val Leu Gln Gln Ile His Pro His Ile Leu Leu Glu Glu Asp Glu
TAT CCC CCA ACC CGA GTT CTC CAA CAG ATA CAC CCG CAC ATA CTG CTG GAA GAA GAC GAA

1081 ————————————————————————————————————————————   ———————————————————
Ile Leu Val Leu Leu Ser Pro Met Thr Ala Tyr Pro Arg Thr Pro Pro Glu Leu Leu Tyr
ATC CTT GTG TTG CTG AGC CCG ATG ACA GCA TAT CCC CGG ACC CCC CCA GAA CTC CTG TAT

1141 ——————————————————————————————*—————————————————————————————————
Pro Glu Ser Asp Gln Asp Gln Leu Glu Pro Leu Glu Glu Glu Glu Glu Glu Tyr Met Pro
CCA GAA AGC GAC CAA GAC CAG CTG GAG CCA CTG GAG GAG GAG GAG GAG GAG TAC ATG CCA
```

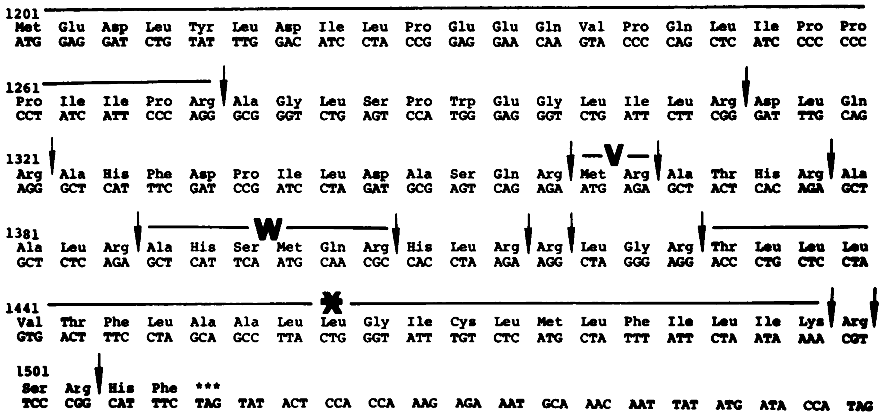

Fig. 6 - The nucleotide sequence of polyoma DNA in the coding regions for the small and medium T antigens, and the predicted amino acid sequence for the two proteins translated in the expected correct reading frames.

the early mRNAs available so far suggests that a splice in the medium T antigen mRNA occurs in this region (R. Kamen, personal communication). The positions of cleavage by trypsin are designated by arrows, termination codons by asterisks and the probable assignments of the methionine-containing tryptic peptides shown in Figure 2 are indicated by lines above the sequences. The reasons for assigning the peptides to the indicated sequences are described below.

Peptide 1 is most likely the N-terminal peptide, N-acetyl-Met-Asp-Arg, for the following reasons: It contains methionine, but not cysteine; it is insensitive to chymotrypsin and V8 protease; its mobility in the electrophoretic dimension is identical to that of the N-terminal peptide of the SV40 large T antigen, N-acetyl-Met-Asp-Lys (which we have identified, using a chemically-synthesized peptide kindly provided by Dr. G. Walter), and its mobility in the chromatographic dimension is slightly faster, as would be predicted for a lysine to arginine change.

Peptide 4 is assigned to nucleotides 275 to 292. It has the correct migration characteristics, does not contain cysteine, and is sensitive to chymotrypsin but not to V8 protease.

Peptide 3 is probably encoded between nucleotides 323 and 376. It is sensitive to both V8 protease and chymotrypsin, and does not contain cysteine.

Peptides 1, 3 and 4 are present in all three T antigens. None of the predicted shared peptides contain cysteine, and none of the observed shared pep-

tides contains cysteine (see Fig. 3).

We can assign the peptides, A, B, C and D, common to the small and medium T antigens, to coding regions between 79 and 85 map units. As shown in Figure 3, peptides B and D contain both methionine and cysteine, and peptides A and C contain methionine, but not cysteine. Based on the predicted and observed mobilities, we would make the following tentative assignments: peptide A (nucleotides 677-766); B (506-535); C (398-430); D (611-676). In addition, there are three predicted peptides in this region which contain cysteine, but not methionine. This number agrees with the observation shown in Figure 3.

We have identified several of the peptides unique to the medium T antigen, demonstrating that they are produced by translation of the polyoma DNA in a different reading frame from that used for the large T antigen in the same region. Two of the peptides, W and V, have been identified by comparison with chemically-synthesized peptides, as noted above, and are assigned to nucleotides 1360-1365 and 1390-1407, respectively. The peptides, U and Y, are assigned as indicated, based on their composition and mobilities. The other predicted peptides, designated by asterisks in Figure 6, cannot be assigned with certainty, although the peptide, nucleotides 1037-1120, may be peptide X.

The failure to detect peptides derived from nucleotides 1021-1275 and nucleotides 1425-1479 may be for a number of reasons. A recurrent problem in analyzing tryptic digests is that the yield of some peptides is non-stoichiometric. This results from the influence of neighboring residues on the efficiency of tryptic cleavage at any given lysine or arginine. For instance Lys-Lys sequences are often not digested to completion. Furthermore, proteins frequently possess trypsin insoluble cores which result in very poor yields of the tryptic peptides from these regions. The peptide coded by nucleotides 1021-1275 is 52 amino acids long and has a large negative charge. Despite the large negative charge we calculate that, because of its mass, this peptide should not be lost from our separation system. We suspect our failure to detect this peptide may be due to the fact that such a big peptide will be rather insoluble in the electrophoresis buffer and will have been lost when the trypsin insoluble material was removed from the digest (6). We are currently looking for this 6,000 dalton peptide by other means including SDS polyacrylamide gel electrophoresis. The methionine and cysteine-containing peptide coded by nucleotides 1425-1497 is extremely hydrophobic. This too we would expect to be very insoluble in electrophoresis buffer. We are trying to detect this peptide by chromatography of the insoluble material removed from the digest.

DISCUSSION

The results described here support and refine the assignment of the coding regions for the polyoma T antigens derived from previous genetic and biochemical evidence. With the DNA sequence of polyoma available, we chose to localize the coding regions for the three polyoma T antigens by correlating the actual polypeptide sequences with those predicted from the DNA sequence. Because of the technical difficulties of direct sequencing, we have used a variety of other techniques to determine the location of several methionine- and cysteine-containing tryptic peptides from the three T antigens on the DNA sequence. The model shown in Figure 7 is the simplest interpretation of our results. The figure shows the three potential reading frames in which polyoma DNA could be translated in the early region. Termination codons are designated by vertical lines and the assignment of the coding regions for the T antigens, derived from analysis of viral mutants and T antigen peptides as described here, are shown. The coding regions for the peptides, V, W, 1 and 14, are indicated.

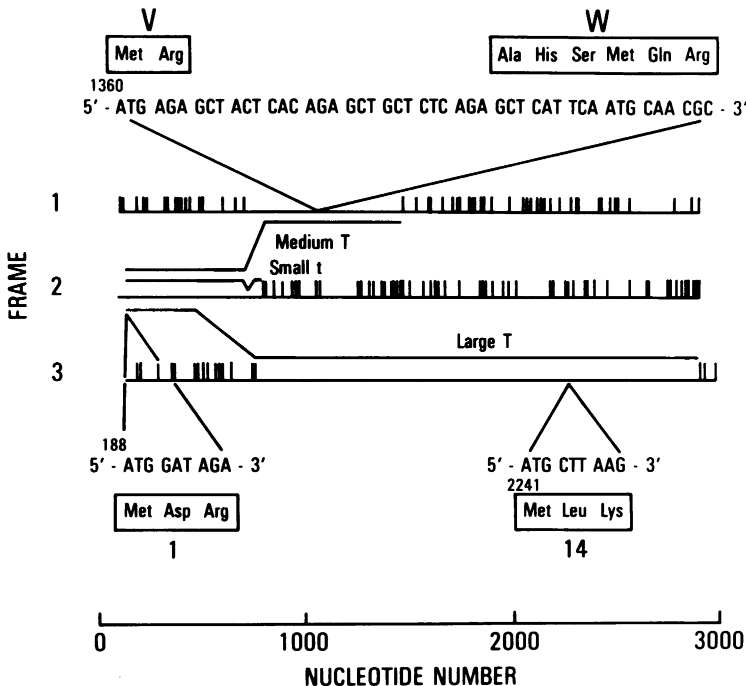

Fig. 7 - Diagram of the three possible reading frames in the early region of polyoma DNA, with coding regions for the three polyoma T antigens indicated.

By in vitro translation of virus-specific RNAs, we have demonstrated that the intervening sequence removed from each of the three early mRNAs must be different and that there must be a splice junction within the translated sequence of each T antigen (1). Hence in our model for the expression of the early region of polyoma, there is a single intervening sequence removed from the mRNA for each T antigen, the sequence removed being different in each case. Since we have not analyzed protein sequences in the region of the putative splice junctions, we have not determined the precise sites where splicing occurs. Furthermore, since we have not correlated the entire protein sequence with the DNA sequence in any case, it would be possible for more complicated splicing patterns to be consistent with our data. However, the analysis of the polyoma early mRNA structures available to date (R. Kamen, personal communication; W. Heiser, personal communication) suggests that the simplest interpretation of our results, as shown in Figure 7, is correct. The exact protein sequences of the three T antigens will most readily be determined by analysis of the sequences of the early mRNAs, as has been done for the SV40 T antigens (20).

The predicted amino acid sequence of the unique region of medium T antigen encoded between 86 and 99 map units has several unusual features. There is a high fraction of hydrophobic residues, including an exclusively hydrophobic region at the C-terminal end (nucleotides 1432 to 1494). There is also a striking organization into acidic and basic domains, with an acidic domain (nucleotides 1069-1236) between two basic domains (nucleotides 958 to 1059 and 1357 to 1512). This region is also rich in proline. Some of these properties may account for the difference between the predicted molecular weight of the medium T antigen (circa 49,000) and the observed molecular weight (60,000) on SDS polyacrylamide gels.

The use of two different reading frames, coding for the large and medium T antigens, over a stretch of approximately 640 nucleotides, is a striking feature of the polyoma genome. The use of overlapping nucleotide sequences, translated in more than one reading frame, is an economical way of encoding proteins using a limited amount of genetic information. This type of organization occurs in the small bacteriophages, ϕX174 and G4 (21,22). The use of overlapping reading frames has also been documented in the late regions of SV40, BK and polyoma. For instance, in SV40 two different reading frames are used over a stretch of 119 nucleotides coding for the N-terminal end of the capsid protein VP1 and the C-terminal ends of VP2 and VP3 (13,14).

Inspection of the DNA sequences of SV40 and polyoma indicates that the

SV40 early region does not share homology with the polyoma early region in the part in which two alternative reading frames are used (86-99 map units)(11). Furthermore, there is only a single open reading frame in the SV40 DNA sequence in the equivalent position (54-40 map units)(13,14). There is, however, a short region between 21 and 16 map units on the SV40 genome, overlapping the C-terminus of the large T antigen, which could code for about 10,000 daltons of protein using an alternate reading frame (13,14). It has been suggested that this might be expressed at some stage in the replicative cycle of SV40 (13,14), possibly through linkage to a polypeptide comprising most of the large T antigen by means of a splice in this region (P. Berg, personal communication). Such a protein might serve the function of a medium T antigen in SV40 (11,12). Alternatively the large T antigen of SV40 might carry out the functions which are allotted to the medium T antigen of polyoma.

The genetic evidence available so far does not allow a clear separation of the functions of the polyoma small and medium T antigens in cell transformation. A more complete description of the polyoma T antigens will probably result from the purification of the individual proteins, and selection of mutants which affect each of the T antigens independently.

ACKNOWLEDGMENTS

This work was supported by Grant Nos. CA 13884, CA 14195, CA 17096 and CA 24288, awarded by the National Cancer Institute, Department of Health, Education and Welfare, and by Grant No. 78SDS, awarded by the Cancer Research Coordinating Committee of the University of California. T. Friedmann was the recipient of a Guggenheim Fellowship. P. Deininger was supported by Training Grant No. GM 07199, awarded by the Nationl Institute of General Medical Sciences, Department of Health, Education and Welfare.

REFERENCES

1. Hunter, T., Hutchinson, M. A., and Eckhart, W. (1978) Proc. Nat. Acad. Sci. U.S.A. 75, 5917-5921.
2. Ito, Y., Spurr, N., and Dulbecco, R. (1977) Proc. Nat. Acad. Sci. U.S.A. 74, 1259-1262.
3. Ito, Y., Brocklehurst, J., and Dulbecco, R. (1977) Proc. Nat. Acad. Sci. U.S.A. 74, 4666-4670.
4. Türler, H., and Salomon, C. (1977) INSERM Colloq. 69, 131-144.
5. Schaffhausen, B., Silver, J., and Benjamin, T. (1978) Proc. Nat. Acad. Sci. U.S.A. 75, 79-83.
6. Hutchinson, M. A., Hunter, T., and Eckhart, W. (1978) Cell 15, 67-77.
7. Fluck, M., Staneloni, R., and Benjamin, T. (1977) Virology 77, 610-624.
8. Eckhart, W. (1977) Virology 77, 589-597.
9. Smart, J. E., and Ito, Y. (1978) Cell 15, 1427-1437.
10. Friedmann, T., Doolittle, R., and Walter, G. (1978) Nature 274, 291-293.

11. Friedmann, T., Esty, A., LaPorte, P., and Deininger, P. (1979) Cell 17, 715-724.
12. Soeda, E., Arrand, J. R., Smolar, N., and Griffin, B. E. (1979) Cell 17, 357-370.
13. Fiers, W., Contreras, R., Haegeman, G., Rogiers, R., Van de Voorde, A., Van Heuverswyn, H., Van Herreweghe, J., Volckaert, G., and Ysebaert, M. (1978) Nature (London) 273, 113-120.
14. Reddy, V. B., Thimmappaya, B., Dhar, R., Subramanian, K. N., Zain, B. S., Pan, J., Celma, C. L., and Weissman, S. M. (1978) Science 200, 494-502.
15. Offord, R. E. (1966) Nature 211, 591-593.
16. Maxam, G., and Gilbert, W. (1977) Proc. Nat. Acad. Sci. U.S.A. 74, 560-564.
17. Sanger, F., Nicklen, S., and Coulson, A. (1977) Proc. Nat. Acad. Sci. U.S.A. 74, 5463-5467.
18. Smith, A. (1979) Nucl. Acids. Res. 6, 831-848.
19. Staden, R. (1977) Nucl. Acids. Res. 4, 4037-4051.
20. Reddy, V. B., Ghosh, P. K., Lebowitz, P., Piatak, M., and Weissman, S. M. (1979) J. Virol. 30, 279-296.
21. Barrell, B. G., Air. G. M., and Hutchison, C. A. (1976) Nature 264, 34-41.
22. Shaw, D. C., Walker, J. E., Northrop, F. D., Barrell, B. G., Gordon, G.N., and Fiddes, J. C. (1978) Nature 272, 510-515.